

# MayLib - a textfile-based bibliographic database for geosciences and a list of references on Devonian matters

Andreas MAY <sup>1</sup>

**Abstract:** Scientists need to manage their own collections of bibliographic data as well as exchange these data easily with colleagues. One solution for this need is MayLib, a bibliographic database that runs on many different operating systems and does not require pre-installation of any software. MayLib is a very small, efficient and comprehensive JAVA program that handles references to publications in any Unicode-compatible language. It is user-friendly and not only contains the basic functions of a bibliographic database, but also has some advanced features. Using MayLib the author has been able to create a list of references, which contains 500 periodicals and 4206 literature references. Of these 4206 references 3191 deal with Devonian matters and 1594 references deal with corals.

**Key Words:** Java; bibliographic database; Devonian; corals; fossils.

**Citation:** MAY A. (2015).- MayLib - a textfile-based bibliographic database for geosciences and a list of references on Devonian matters.- *Carnets Géol.*, Madrid, vol. 15, n° 6, p. 59-62.

**Résumé :** *MayLib - une base "texte" de données bibliographiques pour les géosciences et une liste de références sur des thématiques dévoniennes.*- Tout chercheur a besoin de gérer sa propre collection de données bibliographiques mais aussi d'échanger facilement ces données avec des collègues. MayLib offre une réponse à cette demande : il s'agit d'une base de données bibliographiques, fonctionnant sur plusieurs systèmes d'exploitation et ne nécessitant pas l'installation préalable d'autres logiciels. MayLib est un programme JAVA, léger, complet et efficace, pratique pour manipuler des références à des publications dans toute langue compatible avec le standard Unicode. C'est un outil simple d'utilisation et qui est non seulement doté de toutes les fonctions essentielles d'une base de données bibliographiques mais également de quelques fonctionnalités avancées. Grâce à MayLib, l'auteur a pu constituer une liste de références comportant 500 revues et 4206 entrées. De ces 4206 références, 3191 traitent de thématiques dévoniennes et 1594 ont pour objet des coraux.

**Mots-clefs :** Java ; base de données bibliographiques ; Dévonien ; coraux ; fossiles.

## 1. Introduction

Although big commercial databases and editorials offer an ever increasing amount of bibliographic data, there remains the need to store, organize and query one's own collection of bibliographic data, as well as to interchange these data easily with colleagues. Applications based on relational databases like Microsoft Access® or MySQL® may be very useful for the developers of these applications, nevertheless the use of relational databases makes it difficult to handle the applications and to exchange data easily (For example, a typical problem of applications based on relational databases is that one needs to know how to administer and program relational databases with SQL).

LÖSER (2004) created PaleoTax to record, analyse, and output palaeontological data. A much more flexible answer to the need to manage palaeontological bibliographic data is given by ZALECKA *et al.* (2015). Another answer to this need of geologists, palaeontologists, biologists, and other scientists is MayLib, a

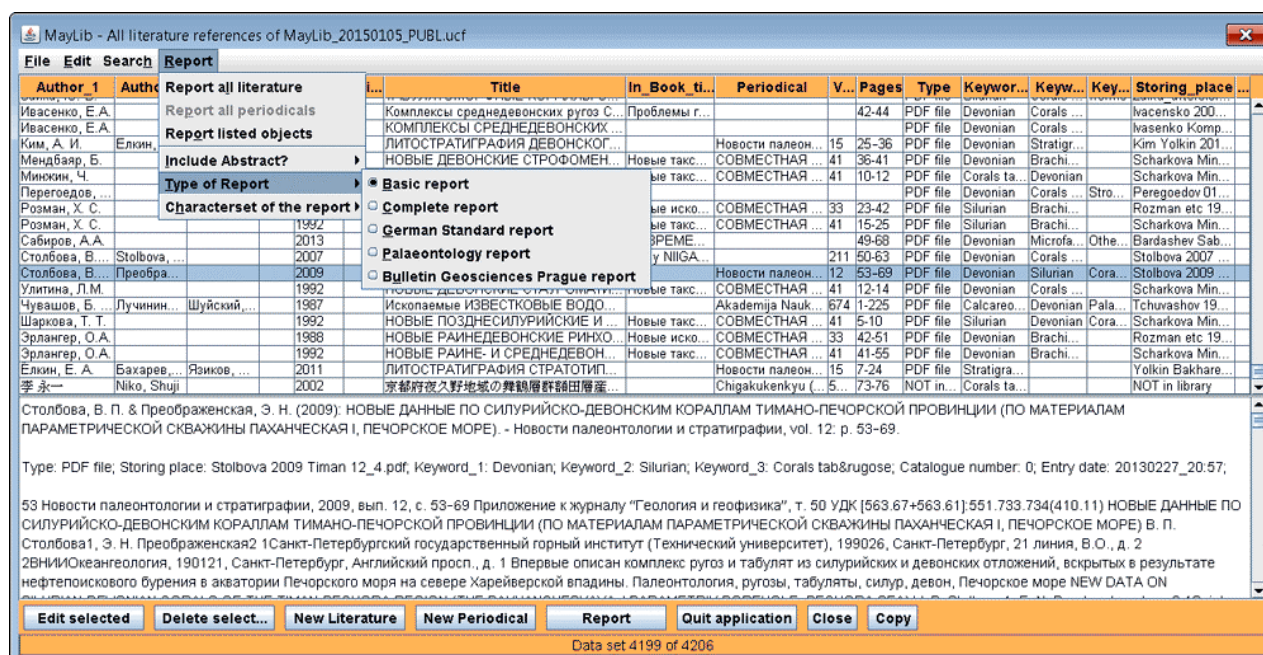
textfile-based bibliographic database, which will be described in this paper. Furthermore, a list of references on Devonian matters and corals, prepared using MayLib, will be presented.

## 2. Features of MayLib

MayLib is a bibliographic database that, in contrast to other databases, does not require a database engine. MayLib needs only a text file with the extension ".UCF" in a specific format. MayLib has been designed to store references to scientific publications (books, articles, theses, etc.) in many languages – English, French, German, Spanish, and Czech as well as Russian, Japanese and Chinese, etc. (see Fig. 1). Because it is programmed in JAVA it runs on many different operating systems (Windows, Mac, Linux, etc.). It does not require pre-installation of any software (except JAVA itself). MayLib is a very small and efficient data management system to handle bibliographic data; it is only 224 KB (!) in size but nevertheless offers comprehensive functionality.

<sup>1</sup> Friedrich-List-Str. 66, D-59425 Unna (Germany)  
[may\\_devonian@yahoo.es](mailto:may_devonian@yahoo.es)

Published online in final form (pdf) on March 9, 2015  
[Editor: Bruno GRANIER; language editor: Robert W. SCOTT]



**Figure 1:** Screenshot of a window of MayLib, showing a search result presented as a table. References to Russian and Japanese publications are visible. The references are ordered alphabetically by authors and year. Furthermore, as can be seen, configuring reports has many possibilities.

MayLib has an intuitive, user-friendly interface. For example, the introduction of new literature references is facilitated by intelligent combo-boxes, checks for duplicate entries and the ability to create new references by copying old ones. Lists of references can be filtered in many ways. The search results can be presented as a table (see Fig. 1) or as a tree (see Fig. 2). Different kinds of reports can be selected (see Fig. 1). The data can be exported to Microsoft Excel®. The operator does not need any programming experience to use MayLib.

MayLib contains all the basic functions of a bibliographic database as defined by FEATHER and STURGES (2003, p. 127) and REITZ (2004, p. 70), but furthermore it also contains several advanced features: As an example, different MayLib text-files can be combined, or a search result saved in another MayLib text-file. Furthermore, for each reference a text of about 40 pages length (up to 95.000 characters) can be stored. These texts can be searched. To make the use of MayLib more comfortable, the "Preferences" menu allows the user to set the default font size.

The text-file which MayLib uses for storing the references has the extension ".UCF", but can be opened with a normal text editor. The format of the UCF file is very similar to the CSV format. To be able to store references in any language, MayLib uses internally, and in the UCF file, a Unicode character set. Nevertheless, MayLib is not only able to read and to write the Unicode character set, but also the default cha-

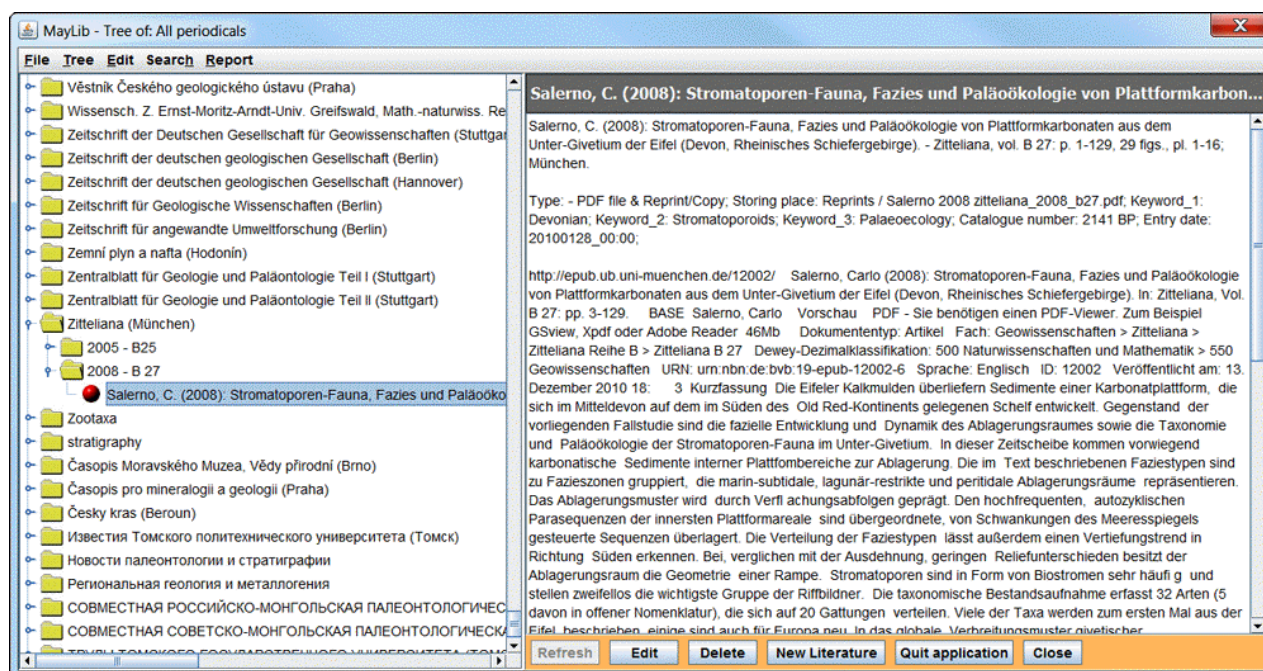
acter set of the operating system of the computer used, as well as the UTF-8 character set. More details on the use of character sets and the data structure of the UCF file are given in the menu item "Mini-Manual" of the "Extras" menu in MavLib.

MayLib has some limitations:

- for each bibliographic reference, only three key words can be defined;
- because of the needs of the UCF file, MayLib converts all semicolons in commas and all line breaks in spaces;
- with a UCF file of about 4.000 references and 35 MB size the performance of MayLib is very fast. However, with a UCF file of more than 50.000 references or more than 300 MB size, its performance may be too slow. Fortunately, MayLib allows you to save search results as new UCF files as well as to combine different UCF files. Therefore it is always possible to avoid oversized UCF files.

### 3. The Programming of MayLib

MayLib has been programmed in JAVA (JavaSE-1.6) using Eclipse®. Consequently, it will run on many different operating systems. The basic elements of the object-oriented design are two classes: the class "Periodical" to store the attributes and methods of the periodicals and the class "Libref" to store the attributes and methods of the literature references.



**Figure 2:** Screenshot of a window of MayLib, showing a list of periodicals presented as a tree. The periodicals are ordered alphabetically by title and place. The subordinate node is year and volume.

Within the UCF file, each line contains all data of a literature reference. By reading the UCF file, MayLib validates the consistency of the data, splits the line into the classes "Periodical" and "Libref" and stores the data in two treemaps, one for each class. The keys of the "Libref" treemap are authors and year of the publication, whereas the keys of the "Periodical" treemap are title and publishing place of the periodical. The use of treemaps, which hold all data ordered in memory, allows a very fast processing of the data. Thanks to the keys of the treemaps, duplicate entries are avoided and all entries are ordered alphabetically.

The data within the treemaps are kept consistent. For example, it is not possible to delete a periodical that is referenced within a literature reference. Every time data is saved, a very large string is synthesized from the two treemaps and written as a complete UCF file to the hard disk. Consequently, the size of the UCF file is critical for the performance of MayLib.

To make a robust graphic user interface, the classes of the SWING toolkit were used. Wherever necessary, derived classes were defined to improve the usability of the elements. Special effort was made to improve the usability of the combo-boxes. Whereas normal SWING combo-boxes only filter by one character, these MayLib combo-boxes can filter by several characters.

#### 4. A list of references on Devonian matters and corals

For the past 21 months the author has been working successfully with the current version of MayLib (Version 1.4., Build 20130325\_18). The

result of this work is the text-file MayLib\_20150105\_PUBL.UCF, which has been assembled from the private library and the literature used by the author. This MayLib text-file from January 2015 is 37 MB in size and contains 500 periodicals and 4206 literature references, many of them with an abstract. Regarding the key words assigned in MayLib, from these 4206 references, 561 references deal with the Silurian, 3191 with the Devonian and 239 with the Carboniferous. 1594 references deal with corals, 483 with stromatoporoids, 372 with brachiopods, and 241 with calcareous algae and microproblematica. Furthermore, 539 references deal with stratigraphy, 428 with reefs and 263 with extinction events. 2918 references exist as PDF files in the author's library.

#### 5. Conclusions

For scientists who want to manage and exchange their own collections of bibliographic data, MayLib is a serious candidate with many advantages. The most important advantages of MayLib are that it is a comprehensive, user-friendly and platform-independent application, which can handle any Unicode-compatible language. Other serious candidates are the JavaScript tools presented by ZALECKA *et al.* (2015). An important advantage of these JavaScript tools is their greater flexibility of data structure, whereas MayLib text-files have a very rigid data structure. PaleoTax cannot be directly compared with these databases, because the program is taxon-based (LÖSER, 2004).

Within the internet several online bibliographic tools are accessible – for more details see



CHILDRESS (2011). Especially important are Zotero (<https://www.zotero.org/>) and Mendeley (<http://www.mendeley.com/>), which are described by PUCKETT (2011) and REISWIG (2010). Both are used normally with a web-browser, but from both websites applications can be downloaded, which are able to work offline. Nevertheless, without connection to the internet, Zotero and Mendeley are restricted in their functionality. In fact, Zotero and Mendeley offer much more possibilities than MayLib to manage many kinds of data sources. The advantages of MayLib are its complete independence from the Internet and its much easier use. Furthermore, users of Zotero and Mendeley run the risks that all users of internet social networks face: hacking of private data, commercialization of initially free services, etc.

Consequently, the reader is invited to compare these applications with each other, and with applications based on relational databases, in order to discern which solution best suits his/her particular needs.

### Acknowledgments

The author acknowledges the support of Peter BRITAIN, who revised the English text.

### Bibliographic references

- CHILDRESS D. (2011).- Citation tools in academic libraries: Best practices for reference and instruction.- *Reference & User Services Quarterly*, Chicago, vol. 51, n° 2, p. 143-152.
- FEATHER J. & STURGES P. (eds., 2003).- International encyclopedia of information and library science.- Routledge, London, Second edition, 720 p.
- LÖSER H. (2004).- PaleoTax - a database program for palaeontological data.- *Computers*

& *Geosciences*, Stanford, vol. 30, p. 513-521. URL with a link to download the software: <http://www.paleotax.de/>

- PUCKETT J. (2011).- Zotero: A guide for librarians, researchers and educators.- The Association of College and Research Libraries, Chicago, 204 p.
- REISWIG J. (2010).- Mendeley.- *Journal of the Medical Library Association*, Chicago, vol. 98, n° 2, p. 193-194.
- REITZ J.M. (2004).- Dictionary for library and information science.- Libraries Unlimited, Westport (Connecticut), 800 p.
- ZAŁECKA K., WRZOLEK T. & GRANIER B. (2015).- Simple and practical techniques to handle small databases, illustrated by a case study: bibliographic data from the "Fossil Cnidaria & Porifera" newsletter (1972-2010).- *Carnets Géol.*, Madrid, vol. 15, n° 2, p. 13-19.

### Appendices

All appendices are licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1) MayLib application (118 KB):

[http://paleopolis.rediris.es/cg/1506/MayLib\\_Application\\_20150105.zip](http://paleopolis.rediris.es/cg/1506/MayLib_Application_20150105.zip)

2) JAVA source code of MayLib (208 KB):

[http://paleopolis.rediris.es/cg/1506/MayLib\\_20130325\\_18\\_Code.zip](http://paleopolis.rediris.es/cg/1506/MayLib_20130325_18_Code.zip)

3) UCF file with references on Devonian matters, corals, etc. (8.033 KB):

[http://paleopolis.rediris.es/cg/1506/MayLib\\_20150105\\_PUBL.zip](http://paleopolis.rediris.es/cg/1506/MayLib_20150105_PUBL.zip)